ESC European Society of Cardiology

European Heart Journal - Digital Health (2025) **6**, 317–325
https://doi.org/10.1093/ehjdh/ztaf019

REVIEW

Explainable artificial intelligence for stroke risk stratification in atrial fibrillation

Raquel Mae Zimmerman [1], Edgar J. Hernandez[2], Martin Tristani-Firouzi[3], Mark Yandell[2], and Benjamin A. Steinberg [4,*]

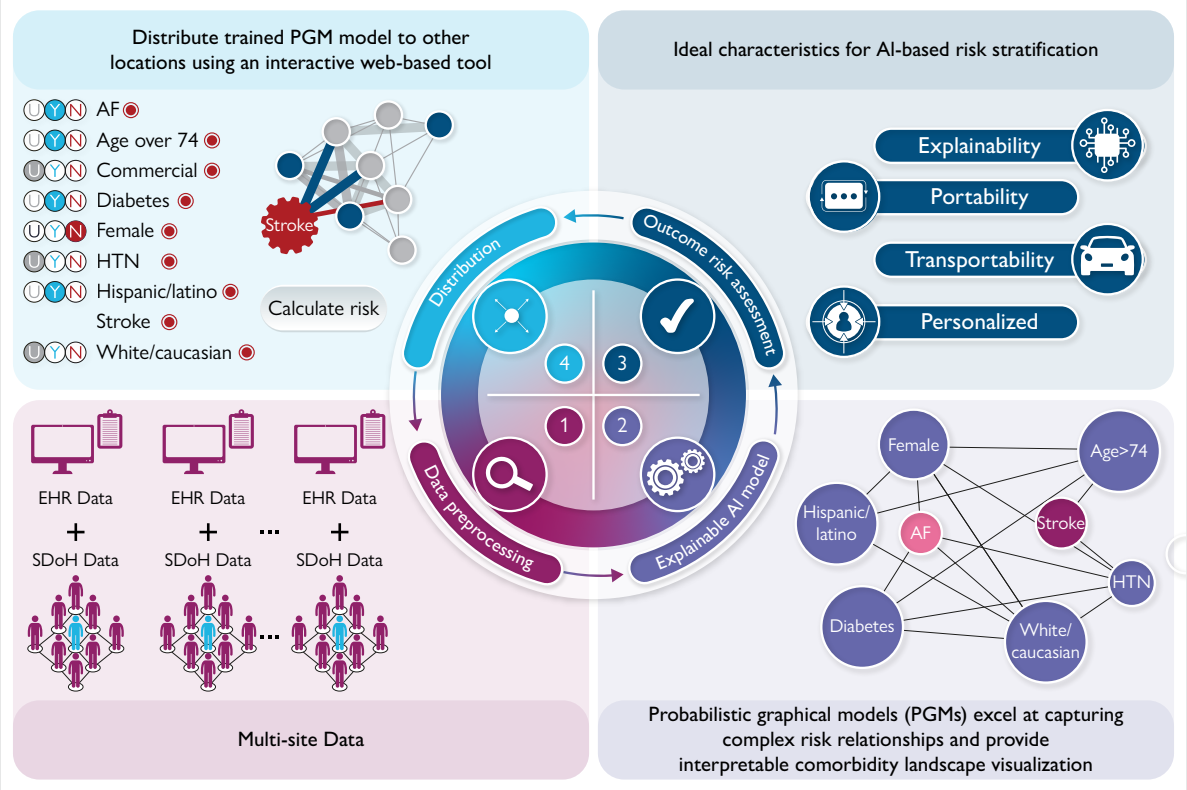
[1]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA; [2]Department of Human Genetics and Utah Center for Genetic Discovery, University of Utah, Salt Lake City, UT, USA; [3]Department of Pediatrics, University of Utah, Salt Lake City, UT, USA; and [4]Department of Medicine, University of Utah, 30 North 1900 East, Room 4A100, Salt Lake City, UT, USA

Received 4 September 2024; revised 17 December 2024; accepted 31 January 2025; online publish-ahead-of-print 22 March 2025

Current risk stratification tools can limit the optimal implementation of new and emerging therapies for patients with heart rhythm disorders. For example, stroke prevention treatments have outpaced means for stroke risk stratification for patients with atrial fibrillation (AF). Artificial intelligence (AI) techniques have shown promise for improving various tasks in cardiovascular medicine. Here, we explain key concepts in AI that are central to using these technologies for better risk stratification, highlighting one approach particularly well suited to the task of portable, personalized risk stratification—probabilistic graphical models (PGMs). Probabilistic graphical models can empower physicians to ask and answer a variety of clinical questions, which we demonstrate using a preliminary model of AF-related stroke risk among 1.6 million patients within the University of Utah Health System. This example also highlights the ability of PGMs to combine social determinants of health and other non-traditional variables with standard clinical and demographic ones to improve personalized risk predictions and address risk factor interactions. When combined with electronic health data, these computational technologies hold great promise to empower personalized, explainable, and equitable risk assessment.

* Corresponding author. Tel: +1 801 581 2121, Email: Benjamin.steinberg@hsc.utah.edu

Graphical Abstract



Keywords Atrial fibrillation • Stroke • Machine learning • Artificial intelligence • Probabilistic graphical model

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia, and its prevalence is rising dramatically.[1–3] Nearly 6 million US adults have AF, and it leads to significant morbidity, mortality, and more than $6B in annual hospitalization costs.[4] The primary source of AF-related morbidity and mortality is stroke, and AF-related strokes are clinically more devastating than strokes of other aetiologies.[5] Systemic oral anticoagulation, with an inherent risk of bleeding, has been the historical standard for stroke prevention in these patients. However, there have been significant improvements in strategies to prevent AF-related strokes and reduce sequelae of anticoagulation (i.e. bleeding). These advances can help fill gaps in treatment among specific patient groups: (i) those miscategorized as 'low stroke risk'; (ii) patients at high risk of stroke who are intolerant of oral anticoagulation; and (iii) patients with stroke events despite appropriate therapy (residual risk). In this setting, novel devices and emerging therapies may be valuable[6,7]; however, limitations in risk stratification preclude appropriate implementation for these patients.

Previous guidelines recommended the CHAD2S2-VASc score for stroke risk stratification among patients with AF—an ordinal score ranging from 0 (lowest risk) to 9 (highest risk), which includes seven risk factors: sex, age, and history of heart failure, vascular disease, hypertension, diabetes, and/or prior stroke.[8,9] However, the newly released European Society of Cardiology (ESC) 2024 guidelines now recommend the CHAD2S2-VA score, which excludes sex as a risk factor, due to its weaker association with stroke risk in broader studies.[1] Despite previous widespread use and endorsements across international guidelines,[9] there are well-known limitations to the CHAD2S2-VASc score: (i) its small derivation population is inherently limiting (1084 patients, only 25 with stroke) and lacks diversity; (ii) some components (e.g. sex) have not held up as robust risk factors in broader studies[10,11]; and (iii) precision of the risk prediction is poor, particularly for patients with low calculated scores.[8,12] Additionally, its ordinal scores cannot account for interactions and dependencies among risk factors, an increasingly recognized phenomenon.[13] Current risk tools are demonstrably poor at distinguishing intermediate risk from truly low stroke risk individuals: up to 7000 strokes still occur annually among ~1 million CHAD2S2-VASc-assessed 'low-risk' patients.[9,14,15] The newly released 2024 ESC guidelines advocate for the use of the CHAD2S2-VA score, which omits the sex category. Although previous guidelines have accounted for gender by setting different thresholds to qualify for oral anticoagulation, excluding gender as a component is a key advance. The updated guidelines formalize this approach, ensuring greater clarity in stroke risk stratification.[16]

Discussion

There remains a critical need for personalized, socially aware, and equitable stroke risk stratification for patients with AF in order to optimally implement contemporary stroke prevention therapies. Explainable artificial intelligence (explainable AI) applications in cardiovascular medicine hold great promise to address such problems of precise risk stratification.[17–20] One specific type of explainable AI tool, the probabilistic

graphical model (PGM), is particularly well suited to this problem. This review outlines the current state of AI techniques for stroke risk stratification for patients with AF, highlights key features for evaluating AI models, provides an overview of PGMs, and uses PGMs to illustrate the promise of explainable AI for stroke risk prediction among patients with AF.

Current state of artificial intelligence and stroke risk stratification in patients with atrial fibrillation

Artificial intelligence models have demonstrated the ability to outperform traditional risk assessment scores, like CHAD2S2-VASc, by combining a wider range of data with more advanced statistical techniques.[21,22] For instance, Lip *et al*. showed that in a large cohort of non-anticoagulated AF and non-AF patients, large improvements in stroke risk prediction can be achieved using logistic regression, with performance metrics significantly higher than that of CHAD2S2-VASc.[23,24] Jung *et al*. and Han *et al*. show similar results with a neural network model and neutral network/random forest model, respectively.[25–27] Both regression and neural networks are established techniques in the field of AI, but reside on opposite ends of the spectrum in terms of complexity and interpretability. While regression is one of the simplest forms of model building, neural networks are composed of complex interconnected layers of artificial neurons allowing for complex, non-linear pattern recognition.

Linear regression provides an interpretable model that assumes a linear relationship between the input features and the target variable. Basic regression models do not capture dependencies between variables without unwieldy interaction terms. Instead, in their simplest form, they assume that the impact of predictors is independent of all others in the absence of multicollinearity. However, in the healthcare domain, diseases often result from a complex interplay of causes, many of which are non-linear and non-independent. *Figure 1* highlights the limitations of such standard approaches, demonstrating that key relationships between CHAD2S2-VASc or CHAD2S2-VA criteria variables cannot be captured by linear, point-based models (see caption for further explanation).

In contrast, neural networks have intricate model architectures and numerous parameters and can learn complex patterns from input data and achieve higher accuracy. However, the complexity of neural network models makes it challenging to interpret how input features contribute to the outcome. Although very accurate, neural nets are often criticized as 'black box' models, and determining how they arrive at risk predictions is difficult. Techniques have been developed to provide insights into neural network decisions, but achieving full transparency remains elusive.[28,29] Additionally, sharing and retraining neural networks in new healthcare systems, particularly in a way that allows for the understanding of differences between sites, presents significant challenges.[30] Other approaches, such as random forest, support vector machine, and K-nearest neighbour, have similar shortcomings regarding complexity and explainability. All of these approaches struggle to capture and quantify multilevel dependencies between variables while maintaining transparency.

Key components for evaluating of artificial intelligence risk stratification tools

Artificial intelligence solutions for stroke risk prediction in AF promise to dramatically improve the quality and equity of care using routinely available electronic health record (EHR) data. However, like every tool, the performance of AI models must be thoroughly evaluated. Recent years have seen many advances in AI models—particularly as computational resources grow almost exponentially and large language models such as ChatGPT become mainstays in research and day-to-day human activity.[31] While many AI models may predict an outcome with high mathematical precision and accuracy, their utility must also be evaluated in other ways. In healthcare specifically, transparency and accountability are crucial to bias-free decision-making. Some of the most important criteria for evaluating the effectiveness of AI models include (i) explainability, (ii) portability, (iii) transportability, and (iv) personalization (see *Table 1*).

Probabilistic graphical model motivations

Numerous complex, interconnected factors, or comorbidities, contribute to an individual's health (or lack thereof). Genetics, environment, lifestyle, and social determinants of health (SDoH) are factors in determining the risk of stroke in patients with AF. Disentangling and distilling these factors into impactful conclusions is a key challenge in the growing field of AI, for which PGMs present a solution.

Probabilistic graphical models model the joint probability distribution of a data set, expressing dependencies between its variables as a graph. Given this graphical model, the Chain Rule of Probability guarantees that we can determine the conditional probability of any combination of variables given the status of any of the other variables (see *Figure 2*).

By way of an example, imagine a data set with just four variables, a, b, c, and d. The data set's full joint probability distribution gives the probability of observing any combination of the four variables. For four binary variables (e.g. true/false), there are 2^4, or 16 combinations. This can be easily represented as a table having 16 rows and 2 columns, one column listing a combination of variables and the other its observed frequency. Allowing each variable to have three states (e.g. true, false, and unknown) can be used to model missing data—a strength of PGMs. With three states per variable, the number of rows in the table grows rapidly, at a rate 3^n. Consider that many variables have more than three states, and many data sets naturally have many variables. Because the size of joint probability table grows exponentially, it is not long before it becomes so large as to be computationally unwieldy.

A key innovation of PGMs lies in their ability to factorize the entire joint distribution of variables into a more compact and structured representation by focusing exclusively on the essential conditional dependencies among variables. These essential dependencies often constitute a smaller subset of all possible combinations of variables, enabling precise inference with minimal computational burden. Moreover, the graphical depiction of the PGM provides a visual roadmap of the essential dependencies among variables, offering researchers a clear and intuitive understanding of the underlying data structure.

Probabilistic graphical models learn the best graphical representation directly from the data using an objective function [such as the Bayesian Information Criterion (BIC)] to discover the optimal model, trading off the number of dependencies (edges) in the model (and hence its size), at the cost of how accurately the model reflects the true joint probability distribution.[35,36] When visualized, the model shows how variables influence and interact. Under the hood, this model structure is also paired with probability distributions that can be used for pairwise or multivariate inferences. The graphical model and corresponding probability distribution tables can be embodied as software and shared quickly and efficiently. Probabilistic graphical models provide intuitive means for illustrating the landscape of dependences between clinical variables and outcomes. Probabilistic graphical model nodes and edges are mapped to real-world medical concepts such as outcomes, comorbidities, and multimorbidities. Directionality in a PGM is crucial to represent the parent–child relationships between the nodes and thus reduces the complexity of the full joint probability distribution

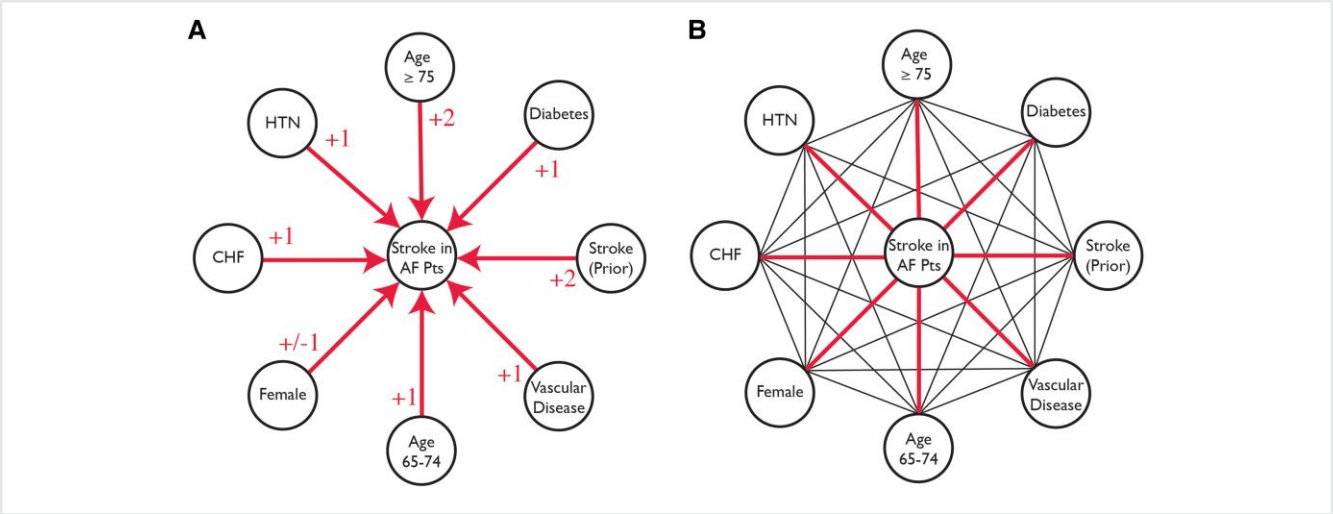


Figure 1 Regression-based models such as CHAD2S2-VASc or CHAD2S2-VA lack the ability to capture complex relationships. Depicted in (*A*) is a graphical model of the original CHAD2S2-VASc scoring system. The graphical representation makes it clear that no interactions between input variables can be captured, resulting in a simple additive model. In the model in (*A*), for example, hypertension and diabetes both contribute one point to a patient's risk. While hypertension and diabetes are individually significant risk factors for stroke, research has consistently demonstrated that they do not contribute to stroke risk equally. Hypertension stands out as the foremost modifiable risk factor for stroke, with estimates suggesting a relative stroke risk of around 4. In contrast, other studies have identified diabetes as a significant stroke risk factor, with relative risks ranging from 1.8 to 3.0. The difference in these estimates is not captured by the point-based system. Moreover, the combined influence of having both conditions simultaneously is entirely overlooked. Understanding these differing contributions and complex interactions of comorbidities is essential for effectively managing and reducing the risk of stroke. Shown on (*B*) is a representation of all possible combinations of the outcome variable (stroke) with other features. As we will discuss, even without the addition of other variables, such as social determinants of health variables, some of these connections among variables are significant, and their interdependencies need to be captured. Probabilistic graphical models address this need. HTN, hypertension; CHF, congestive heart failure.

Table 1 Key concepts in the evaluation of artificial intelligence models

Explainability	Explainable models provide clear, understandable insights into their decision-making processes, making their outcomes interpretable. They also prioritize transparency, allowing users to understand the reasoning behind predictions. Explainable models are relatable to a benchmark/gold standard truth data set. In the case of stroke and AF, for example, the current benchmark is the CHAD2S2-VA scoring system
Portability	The ability for models to be transported between EHR systems for analysis and comparison on various data sets using standard data formats as input. This requires operating system, environment, and hardware compatibility
Transportability	The ability for a model to work 'out of the box' without retraining for clinical decision support or risk assessment at an external site. Such a model will perform well on many external data sets—i.e. it is generalizable. This yields personalization not just by individual, but at the system/local population level
Personalization	The ability to use the model to tailor treatment plans or health recommendations to individual patients based on their personal medical history, genetics, and/or lifestyle factors.[32] Personalization highlights a shift from a traditional 'one size fits all' approach seen in the CHAD2S2-VA and CHAD2S2-VASc scoring system to more precise risk calculations and subsequent treatment decisions for individual patients
Equitability	Equitability ensures that AI models deliver fair and unbiased performance across diverse patient populations, understanding differences in demographics, socio-economic factors, and comorbidities. This is critical for reducing healthcare disparities and ensuring that predictions are equally reliable for all groups[22]
Performance	Performance refers to the overall accuracy, reliability, and efficiency of the model, as evaluated by key metrics such as sensitivity, specificity, precision, and area under the curve (AUC). In stroke risk prediction, for instance, high performance means the model successfully distinguishes high-risk patients from low-risk individuals, outperforming existing tools like CHAD2S2 while maintaining consistency across populations[33]

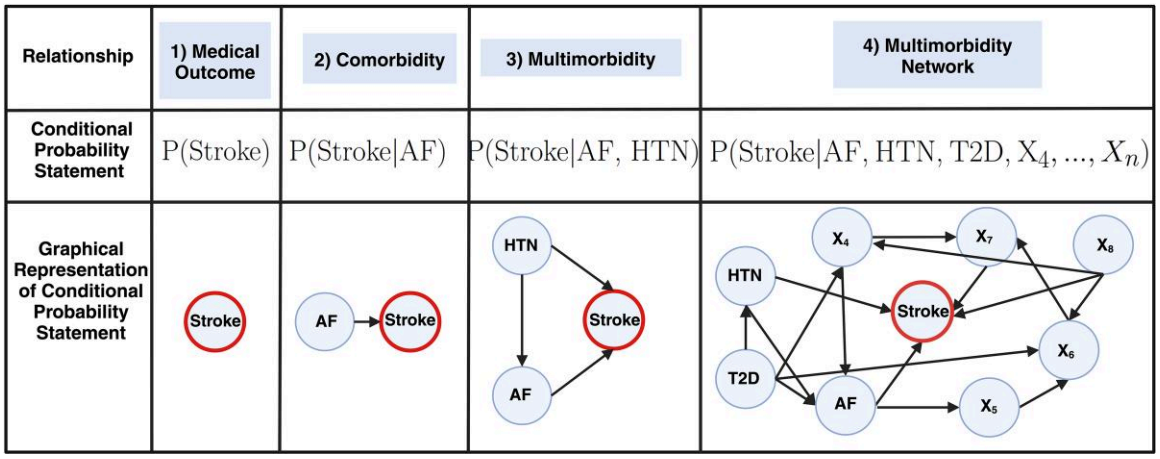


Figure 2 Definitions of relationships between medical terms with corresponding probability statements and graphical representations.[34] Probabilistic graphical models can be used to tease apart the intertwined, conditionally dependent impacts of variables across domains using this notation. The population prevalence of a medical outcome is the starting point for all risk estimations, for example, the frequency of stroke in a population P(stroke). Second, comorbid risk estimates restrict this calculation to a subset of the population with a single indication, for example, those with atrial fibrillation (AF), the main risk factor for stroke. More formally, a comorbid risk estimate can be written as a conditional probability statement, e.g. P(Stroke | AF), the probability of stroke given atrial fibrillation. Conditional probability statements can also be represented using graphical notation, as is shown in Row 3. In this notation, the circles or 'nodes' represent the variables of interest and a line or 'edge' drawn between them denotes that the variables are not independent of one another but are conditionally dependent. If stroke and atrial fibrillation were independent of one another, there would be no line or 'edge' between the two variables in the graphical representation. This is rarely the case, as clinical variables are often conditionally dependent. Indeed, dealing with this real-world complication is a primary motivation for using probabilistic graphical models, especially for *multimorbid* risk estimates, which ask questions about morbidity in the context of multiple indications. For example, third, the P(Stroke | AF, HTN) represents the probability of a patient having a stroke if they also have atrial fibrillation and hypertension. Here, the graphic representation offers a clear advantage over the conventional notation, as it provides means to denote that none of the indications are conditionally independent of the outcome, stroke. Moreover, hypertension and atrial fibrillation are also correlated with one other. Finally, we see a multimorbidity network, with the conditional landscape of numerous variables. Traditional notation struggles to capture the dependency shown. The graphical notation shown in Row 3 better represents the true relationships of comorbidities and provides a lightweight way to share the joint probability distribution. T2D, type 2 diabetes mellitus.

of the model. In many applications, particularly in the biological sciences, this directionality is not assumed to be associated with causality but rather serves as a convenient way to express observed dependencies. In the most widely used type of PGM, Bayesian networks, the model is represented formally as a directed acyclic graph, meaning that the edges (arrows) between nodes (variables) form a directed graph without any cycles, thus eliminating the possibility of feedback loops and absorbing states. *Figure 2* provides a summary example of the relationships between medical terms with corresponding conditional probability statements and graphical representations.

The need to model complex multimorbid outcomes, as depicted in *Figure 2*, is a primary motivation for using PGMs in medicine. Moreover, PGMs also meet all of the criteria discussed in *Table 1* (explainable, transportable, portable, and personalized), making them ideal for applications such as stroke risk stratification among patients with AF.

Probabilistic graphical models are *fully transparent*, meaning any factor's or combination of factors' absolute contribution to risk is correctly quantified and easily obtained.[37–39] Additionally, once derived, a PGM can compute the probability of any of its variables rather than deriving a new model for a new outcome—a significant advantage compared to techniques, such as multivariate regression, wherein a new model must be created for every outcome. For example, with a PGM, calculating the risk of stroke in an 80-year-old patient with AF and vascular disease is as manageable as calculating the probability of heart failure in an 80-year-old patient with AF who undergoes catheter ablation for AF. Only a single model is needed. Probabilistic graphical models address this need. Probabilistic graphical models are also t*ransportable*—they can then be shared without the exchange of protected health information and enable fast and flexible risk assessments.[40] These features of PGMs, combined with their graphical representations, can be used to elucidate and visualize the 'risk landscape' of stroke and AF. Helping to build intuition and understanding of how biases and healthcare disparities impact risk.[40–43]

Preparing the data

The use of PGMs in conjunction with large EHR data sets, supplemented with SDoH information, has great potential to improve the strength and equity of risk stratification in patients with AF. Creating a PGM begins with identifying a clinical question and a relevant data set. The EHR data are an obvious resource for PGM creation, containing diagnoses, procedures, lab tests, billing, and other medication information, all described using standardized medical terminologies. This information can be further supplemented using natural language processing technologies to extract additional information from clinical notes.[41] The SDoH data can also be incorporated, enabling more precision and personalization. This includes information on education, income, employment status, housing conditions, and access to healthcare resources.

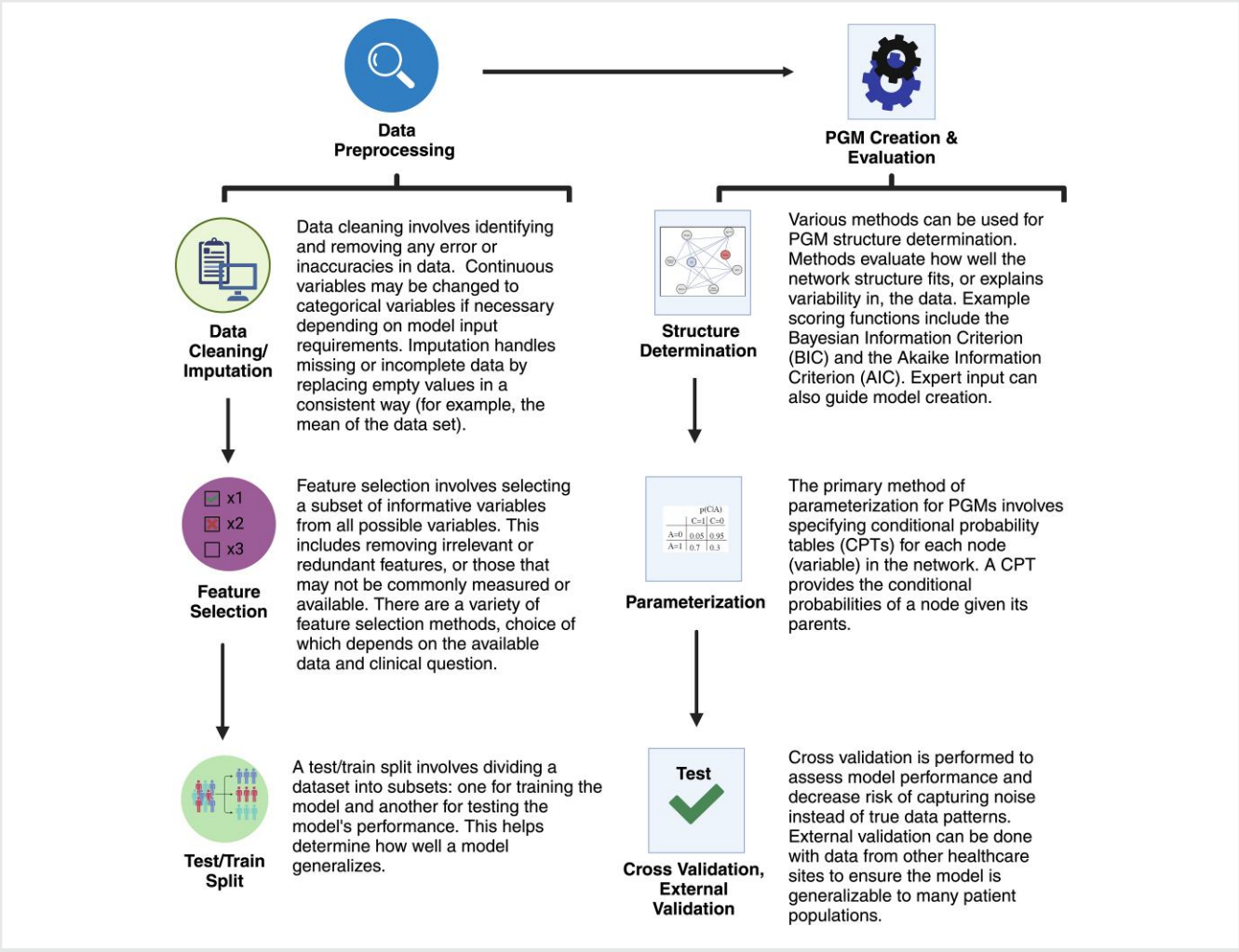


Figure 3 Overview of probabilistic graphical model creation, including data pre-processing and model creation and evaluation. Data pre-processing involves data cleaning and imputation, feature selection, and performing a test/train split. Probabilistic graphical model creation and evaluation require structure determination, parameterization, and cross and external validation.[34]

Incorporating these diverse data sources can provide a holistic understanding of the factors influencing stroke in patients with AF.

Once data are selected, it must be processed, as detailed in *Figure 3*. This processing includes data cleaning and imputation, feature selection, and splitting the data into test/train sets. Data cleaning involves identifying and removing formatting differences and obvious inaccuracies or errors in data. While there is no formal requirement excluding the use of continuous variables in PGM construction, for practical applications using finite data sets, continuous variables often need to be discretized. These continuous variables can be transformed to categorical ones using statistical methods (quantiles, standard deviations from the mean), or by considering domain-specific knowledge (accepted ranges for low, normal, and high lab values, for instance). Imputation can be used to 'fill in' missing or incomplete data by replacing empty values with substitutes in a consistent manner (for example, the mean value of the variable for the data set), or the 'missingness' can be directly incorporated into the variable's definition, e.g. true, false, or unknown. Imputation is an optional step as most PGM-building algorithms will allow for missing data.

Feature selection involves choosing a subset of informative variables from all possible ones. This includes removing irrelevant features and redundant features, or those that may not be commonly measured or available. There are a variety of feature selection methods, and the choice of feature selection method depends on the available data and clinical questions.[40,41]

Finally, a test/train split involves dividing a data set into subsets: one for training the model and another for testing the model's performance. Using a test/train split provides a better estimate of how well the model generalizes and aims to ensure it performs well on diverse data. The training data are used for determining the structure of the model and its parameters.

Creating a probabilistic graphical model

Multiple approaches to PGM structure determination and parametrization exist, but all aim to maximize how well the network structure fits the data. This is accomplished through the use of a scoring function such as the BIC, which penalizes model complexity, all else being equal, favouring a model with fewer edges. Expert input can also guide model creation, for instance, prohibiting connections that are impossible in the real world (blacklisting) or including connections between nodes that should always be present in the graph

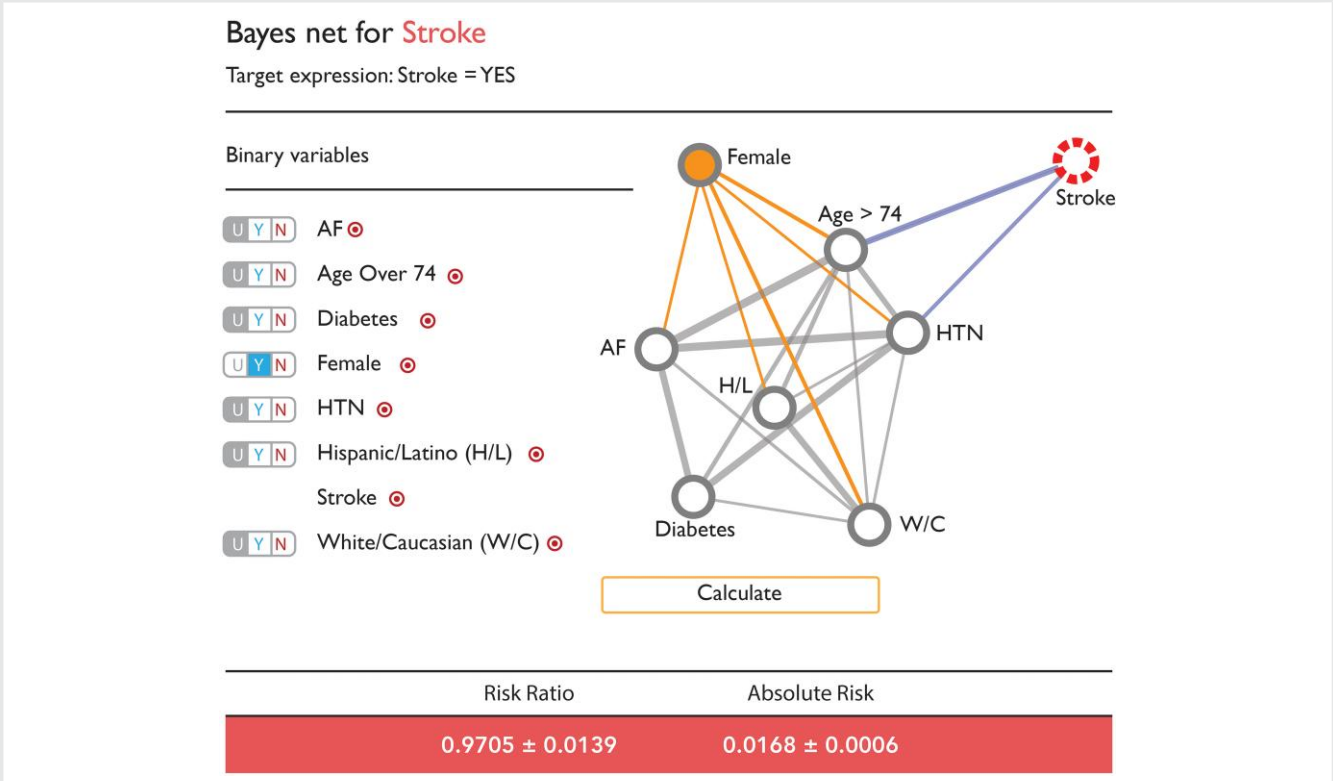


Figure 4 Proof-of-principle probabilistic graphical model web-based interface. This probabilistic graphical model was trained on 100 000 randomly selected UHealth electronic health records. In this example, the projected risk of stroke is shown for a female patient with atrial fibrillation and no other known comorbidities. In this cohort, female gender alone does not convey additional risk. This is consistent with the previous studies which have emphasized that being female alone does not inherently increase the risk of certain conditions but rather acts as a risk modifier in the context of other comorbidities, e.g. hypertension, age, ancestry, and ethnicity. Y, factor present; N, factor not present; U, factor status unknown; HTN, hypertension.

structure (whitelisting). The primary method of parameterization involves specifying conditional probability tables (CPTs) for each node in the network. A CPT provides the conditional probabilities of a node given its parents. The values in the CPT can be identified with either maximum likelihood estimation (AIE) or Bayesian estimation. The AIE finds the values that maximize the likelihood of observing the given data under the model. Bayesian estimation incorporates prior beliefs or knowledge about parameters, updating them with observed data to obtain the best fit.[44] Cross-validation is then used to assess model performance and decrease the risk over fitting the model to the training data. External validation can also be performed with data from other sites to ensure the model is generalizable to many patient populations.

The result is a graphical and computational model of the joint probability distribution and the dependencies among its variables. As explained in the Introduction, the joint probability distribution specifies the probabilities associated with all possible combinations of variable assignments, a crucial step in modelling complex relationships and facilitating inference. An example of such a network model is the multimorbidity network for stroke patients shown in the Graphical Abstract. This graphical model and its corresponding probability distribution tables can be quickly and easily shared. The model can also be easily retrained as more data become available. Moreover, model performance can be continually evaluated and improved.

Probabilistic graphical model transportability

As mentioned previously, a major advantage of PGMs is their transportability. Probabilistic graphical models can be shared as web-based tools, allowing users to independently explore the outcomes landscape (aka the joint probability distribution). Building upon previous results in other healthcare domains,[40,41,45] we recently published a PGM for forecasting stroke among patients with AF using 1.6 million patients within the University of Utah Health System.[45,46] A PGM example trained on 100 000 University of Utah Health (UHealth) patients is shown in *Figure 4*. Note the dependencies among the variables, such as that between female gender, advanced age, and hypertension. Here, we observe that gender is important in the context of other comorbidities but not in isolation. This has been highlighted by previous studies showing that female gender alone does not convey increased risk but is rather a risk modifier in the context of other comorbidities.[10] The ability of PGMs to capture relations such as these is an essential prerequisite for improving risk stratification. For more on this point, see the discussion of *Figure 1*. The PGM-based tools like the one illustrated in *Figure 4* can also serve as the primary intervention for a clinical trial of alternative risk stratification strategies—e.g. patient- or site-level randomization to standard risk stratification (i.e. CHAD2S2-VA) vs. PGM-guided risk stratification. In essence, each net serves as a 'working prototype' of an electronic decision support tool, thereby

laying the foundation for ultimate deployment with Fast Healthcare Interoperability Resources for direct accessibility through the EHR.[47,48]

Conclusions and future directions

Stroke remains the primary contributor to morbidity and mortality in patients with AF despite notable advancements in prevention. While effective stroke prevention strategies are available, their optimal utilization is hindered today by rudimentary stroke risk stratification tools such as CHAD2S2-VASc and CHAD2S2-VA and by disparities in both healthcare access and AF-related outcomes. Current AI-based risk assessments lack the precision and awareness required to stratify patients optimally—for example, patients may be incorrectly classified as low risk but have other comorbidities or risk factors that interact to elevate stroke risk. *Figure 1* highlights the limitations of current standard approaches, demonstrating that key interactions between CHAD2S2-VA criteria variables are not captured and that a point-based scoring criterion may not be the best way to quantify patients' risk. In contrast, PGMs excel at capturing complex risk relationships that arise from comorbidity interactions. Moreover, PGMs provide for interpretable comorbidity landscape visualization, can be shared with other sites, and can be easily retrained to apply the model to diverse patient populations. For patients with AF, these models have the potential to not only aid in stroke prevention but also help select patients for aggressive rhythm control and/or candidacy for catheter ablation with respect to any number of outcomes. More broadly, risk stratification for sudden cardiac death represents a major challenge across numerous disease states, with some approaches limited by very small populations (e.g. inherited arrhythmia disorders). The methods described here can be used to elucidate the risk conferred in numerous contexts, even including social determinants of health or other contextual modifiers. Probabilistic graphical models offer the opportunity to truly leverage large EHR data for the improved personalization of care through more precise, portable risk stratification—a pivotal task among patients with heart rhythm disorders.

Author contributions

Raquel Mae Zimmerman (Writing—original draft preparation [lead], Data curation, Formal Analysis, Visualization, Writing—review & editing), Edgar J. Hernandez (Data curation, Methodology, Software, Visualization, Writing—review & editing), Martin Tristani-Firouzi (Conceptualization, Funding acquisition, Methodology, Supervision, Writing—review & editing), Mark Yandell (Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Resources, Software, Supervision, Writing—review & editing), and Benjamin A. Steinberg (Conceptualization, Funding acquisition, Supervision, Writing—review & editing).

Funding

This work was supported by grants from the National Institutes of Health to BAS (#R56HL168264, #1R01HL177105) and RMZ (#F30HL17068901). This research was also supported by the National Institutes of Health, Interdisciplinary Training Grant T32 Program in Computational Approaches to Diabetes and Metabolism Research, #1T32DK11096601. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest: My reports being a co-founder of and consultant to BackDrop Health, Inc. and Fabric Genomics. BAS reports research grants or contracts from Biosense Webster, Boston Scientific, sanofi, and AltaThera; consulting to Milestone, Sanofi, and AltaThera; lecture fees from AltaThera; DSMB activities for Element Science, and stock ownership in Doximity. No other conflicts were reported.

Data availability

No new data were generated or analysed in support of this research.

Lead author biography

Raquel Mae Zimmerman is a fourth-year MD-PhD student with a passion for data-driven, innovative approaches to precision medicine. Her research focuses on leveraging artificial intelligence and machine learning to improve risk prediction models and address disparities in healthcare. She aims to develop innovative tools to assist clinicians in making personalized, evidence-based decisions for improved patient outcomes. Outside of research, she enjoys running, hiking, and organizing community wellness initiatives. She hopes to pursue a career as a physician-scientist in oncology.

References

1. Van Gelder IC, Rienstra M, Bunting KV, Casado-Arroyo R, Caso V, Crijns HJGM, *et al.* 2024 ESC guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): developed by the task force for the management of atrial fibrillation of the European Society of Cardiology (ESC), with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Endorsed by the European Stroke Organisation (ESO). *Eur Heart J* 2024;**45**:3314–3414.
2. Ma Q, Zhu J, Zheng P, Zhang J, Xia X, Zhao Y, *et al.* Global burden of atrial fibrillation/flutter: trends from 1990 to 2019 and projections until 2044. *Heliyon* 2024;**10**:e24052.
3. Liu Y-M, Wang W, Zhang X, Lei F, Qin J-J, Huang X, *et al.* Frontiers | the rising death burden of atrial fibrillation and flutter in low-income regions and younger populations. *Front Epidemiol* 2023;**3**:1122790.
4. Wolowacz SE, Samuel M, Brennan VK, Jasso-Mosqueda J-G, Van Gelder IC. The cost of illness of atrial fibrillation: a systematic review of the recent literature. *Europace* 2011;**13**:1375–1385.
5. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, *et al.* Heart disease and stroke statistics-2021 update: a report from the American Heart Association. *Circulation* 2021;**143**:e254–e743.
6. Freeman JV, Varosy P, Price MJ, Slotwiner D, Kusumoto FM, Rammohan C, *et al.* The NCDR left atrial appendage occlusion registry. *J Am Coll Cardiol* 2020;**75**:1503–1518.
7. Reddy VY, Doshi SK, Sievert H, Buchbinder M, Neuzil P, Huber K, *et al.* Percutaneous left atrial appendage closure for stroke prophylaxis in patients with atrial fibrillation: 2.3-year follow-up of the PROTECT AF (watchman left atrial appendage system for embolic protection in patients with atrial fibrillation) trial. *Circulation* 2013;**127**:720–729.
8. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC Jr, *et al.* 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol* 2019;**74**:104–132.
9. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomstrom-Lundqvist C, *et al.* 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the task force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J* 2021;**42**:373–498.
10. Nielsen PB, Skjoth F, Overvad TF, Larsen TB, Lip GYH. Female sex is a risk modifier rather than a risk factor for stroke in atrial fibrillation: should we use a CHA(2)DS(2)-VA score rather than CHA(2)DS(2)-VASc? *Circulation* 2018;**137**:832–840.
11. Mikkelsen AP, Lindhardsen J, Lip GYH, Gislason GH, Torp-Pedersen C, Olesen JB. Female sex as a risk factor for stroke in atrial fibrillation: a nationwide cohort study. *J Thromb Haemost* 2012;**10**:1745–1751.
12. Friberg L, Benson L, Rosenqvist M, Lip GYH. Assessment of female sex as a risk factor in atrial fibrillation in Sweden: nationwide retrospective cohort study. *Br Med J* 2012;**344**:e3522.

13. Buhari H, Fang J, Han L, Austin PC, Dorian P, Jackevicius CA, *et al.* Stroke risk in women with atrial fibrillation. *Eur Heart J* 2024;**45**:104–113.
14. McIntyre WF, Linz D. Atrial fibrillation and stroke: who is low risk and what are we going to do about it? *Eur Heart J* 2022;**43**:3539–3541.
15. Komen JJ, Pottegard A, Mantel-Teeuwisse AK, Forslund T, Hjemdahl P, Wettermark B, *et al.* Oral anticoagulants in patients with atrial fibrillation at low stroke risk: a multicentre observational study. *Eur Heart J* 2022;**43**:3528–3538.
16. Boriani G, Vitolo M, Mei DA. CHA2DS2-VA instead of CHA2DS2-VASc for stroke risk stratification in patients with atrial fibrillation: not just a matter of sex. *EP Europace* 2024; **26**:euae281.
17. Petzl AM, Jabbour G, Cadrin-Tourigny J, Pürerfellner H, Macle L, Khairy P, *et al.* Innovative approaches to atrial fibrillation prediction: should polygenic scores and machine learning be implemented in clinical practice? *EP Europace* 2024;**26**:euae201.
18. Lolak S, Attia J, McKay GJ, Thakkinstian A. Comparing explainable machine learning approaches with traditional statistical methods for evaluating stroke risk models: retrospective cohort study. *JMIR Cardio* 2023;**7**:e47736.
19. Gavidia M, Zhu H, Montanari AN, Fuentes J, Cheng C, Dubner S, *et al.* Early warning of atrial fibrillation using deep learning. *Patterns* 2024;**5**:100970.
20. Lu J, Bisson A, Bennamoun M, Zheng Y, Sanfilippo FM, Hung J, *et al.* Predicting multifaceted risks using machine learning in atrial fibrillation: insights from GLORIA-AF study. *Eur Heart J Digit Health* 2024;**5**:235–246.
21. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina (Kaunas)* 2020;**56**:455.
22. van Smeden M, Heinze G, Van Calster B, Asselbergs FW, Vardas PE, Bruining N, *et al.* Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease. *Eur Heart J* 2022;**43**:2921–2930.
23. Lip GYH, Tran G, Genaidy A, Marroquin P, Estes C, Landsheft J. Improving dynamic stroke risk prediction in non-anticoagulated patients with and without atrial fibrillation: comparing common clinical risk scores and machine learning algorithms. *Eur Heart J Qual Care Clin Outcomes* 2022;**8**:548–556.
24. Kalarus Z, Mairesse GH, Sokal A, Boriani G, Średniawa B, Casado-Arroyo R, *et al.* Searching for atrial fibrillation: looking harder, looking longer, and in increasingly sophisticated ways. An EHRA position paper. *EP Europace* 2023;**25**:185–198.
25. Jung S, Song M-K, Lee E, Bae S, Kim Y-Y, Lee D, *et al.* Predicting ischemic stroke in patients with atrial fibrillation using machine learning. *Front Biosci (Landmark Ed)* 2022; **27**:80.
26. Han L, Askari M, Altman RB, Schmitt SK, Fan J, Bentley JP, *et al.* Atrial fibrillation burden signature and near-term prediction of stroke: a machine learning analysis. *Circ Cardiovasc Qual Outcomes* 2019;**12**:e005595.
27. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform* 2021;**113**:103655.
28. Saleem R, Yuan B, Kurugollu F, Anjum A, Liu L. Explaining deep neural networks: a survey on the global interpretation methods. *Neurocomputing (Amst)* 2022;**513**:165–180.
29. Molnar C. Interpretable machine learning. In: *A Guide for Making Black Box Models Explainable*. Independently Published; 2023.
30. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: a scoping review. *PLoS One* 2019;**14**:e0212356.
31. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med* 2023;**29**:1930–1940.
32. Hassan M, Awan FM, Naz A, deAndrés-Galiana EJ, Alvarez O, Cernea A, *et al.* Innovations in genomics and big data analytics for personalized medicine and health care: a review. *Int J Mol Sci* 2022;**23**:4645.
33. Al-Zaiti SS, Alghwiri AA, Hu X, Clermont G, Peace A, Macfarlane P, *et al.* A clinician's guide to understanding and critically appraising machine learning studies: a checklist for ruling out bias using standard tools in machine learning (ROBUST-ML). *Eur Heart J Digit Health* 2022;**3**:125–140.
34. Pintaric S. Figure Created with BioRender. BioRender; 2023.
35. Kuha J. AIC and BIC. *Sociol Methods Res* 2016;**33**:188–229.
36. Zhang D, Chen M-H, Ibrahim JG, Boye ME, Wang P, Shen W. Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials. *Stat Med* 2014;**33**:4715–4733.
37. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Cent Rep* 2019;**49**:15–21.
38. Heckerman D, Geiger D, Chickering DM. Learning bayesian networks: the combination of knowledge and statistical data. *arXiv preprint* 2013. doi: https://doi.org/10.48550/arXiv.1302.6815
39. Leopold F, Shrestha YR, Paudel B. A deep learning pipeline for patient diagnosis prediction using electronic health records. *arXiv preprint 2020:arXiv:2006.16926*. 2020. doi: https://doi.org/10.48550/arXiv.2006.16926
40. Wesolowski S, Lemmon G, Hernandez EJ, Henrie A, Miller TA, Weyhrauch D, *et al.* An explainable artificial intelligence approach for predicting cardiovascular outcomes using electronic health records. *Public Lib Sci Digit Health* 2022;**1**:e0000004.
41. Lemmon G, Wesolowski S, Henrie A, Tristani-Firouzi M, Yandell M. A poisson binomial-based statistical testing framework for comorbidity discovery across electronic health record datasets. *Nat Comput Sci* 2021;**1**:694–702.
42. Watkins WS, Hernandez EJ, Wesolowski S, Bisgrove BW, Sunderland RT, Lin E, *et al.* De novo and recessive forms of congenital heart disease have distinct genetic and phenotypic landscapes. *Nat Commun* 2019;**10**:4722.
43. Miller TA, Hernandez EJ, Gaynor JW, Russell MW, Newburger JW, Chung W, *et al.* Genetic and clinical variables act synergistically to impact neurodevelopmental outcomes in children with single ventricle heart disease. *Commun Med (Lond)* 2023;**3**:127.
44. Zyphur MJ, Oswald FL. Bayesian estimation and inference. *J Manage* 2013;**41**:390–420.
45. Zimmerman RM, Hernandez EJ, Watkins WS, Blue N, Tristani-Firouzi M, Yandell M, *et al.* An explainable artificial intelligence approach for discovering social determinants of health and risk interactions for stroke in patients with atrial fibrillation. *Am J Cardiol* 2023;**201**:224–226.
46. Zimmerman RM, Hernandez EJ, Watkins WS, Blue N, Tristani-Firouzi M, Yandell M, *et al.* An explainable AI approach for discovering social determinants of health and risk interactions for stroke in patients with atrial fibrillation. *medRxiv* 2023. doi: https://doi.org/10.1101/2023.02.27.23286470
47. Taber P, Radloff C, Del Fiol G, Staes C, Kawamoto K. New standards for clinical decision support: a survey of the state of implementation. *Med Inform (Lond)* 2021;**30**:159–171.
48. Kawamoto K, Kukhareva PV, Weir C, Flynn MC, Nanjo CJ, Martin DK, *et al.* Establishing a multidisciplinary initiative for interoperable electronic health record innovations at an academic medical center. *JAMIA Open* 2021;**4**:ooab041.